An Aggregation Approach to Short-Term Traffic Flow Prediction

Man-Chun Tan, S. C. Wong, Jian-Min Xu, Zhan-Rong Guan, and Peng Zhang

***Abstract*—In this paper, an aggregation approach is proposed for traffic flow prediction that is based on the moving average (MA), exponential smoothing (ES), autoregressive MA (ARIMA), and neural network (NN) models. The aggregation approach assembles information from relevant time series. The source time series is the traffic flow volume that is collected 24 h/day over several years. The three relevant time series are a weekly similarity time series, a daily similarity time series, and an hourly time series, which can be directly generated from the source time series. The MA, ES, and ARIMA models are selected to give predictions of the three relevant time series. The predictions that result from the different models are used as the basis of the NN in the aggregation stage. The output of the trained NN serves as the final prediction. To assess the performance of the different models, the naïve, ARIMA, nonparametric regression, NN, and data aggregation (DA) models are applied to the prediction of a real vehicle traffic flow, from which data have been collected at a data-collection point that is located on National Highway 107, Guangzhou, Guangdong, China. The outcome suggests that the DA model obtains a more accurate forecast than any individual model alone. The aggregation strategy can offer substantial benefits in terms of improving operational forecasting.**

***Index Terms*—Autoregressive moving average (ARIMA) model, data aggregation (DA), exponential smoothing (ES), moving average (MA), neural network (NN), time series, traffic flow prediction.**

I. Introduction

TRAFFIC flow forecasting is an essential part of transportation planning, traffic control, and intelligent transportation systems [1]–[19]. In particular, short-term traffic volume forecasts support proactive dynamic traffic control. As a result, forecasting technologies have gotten the attention of traffic engineers and researchers. A wide variety of techniques has been applied in the context of short-term traffic flow forecasting, depending upon the type of data that are available and the potential end use of the forecast. These techniques include time series analysis [6], [7], Bayesian networks [8], neural networks (NNs) [9]–[12], fuzzy NNs [13], [14], nonparametric regression (NP) [15], [16], and intelligence computation [17]–[19].

It is almost universally agreed in the forecasting literature that no single method is the best in every situation. Since the early work of Edgerton and Kolbe [20] and Bates and Granger [21], the literature on this topic has significantly expanded. Numerous researchers have demonstrated that combining the predictions of several models frequently results in prediction accuracy higher than that of the individual models [22], [23]. Using a hybrid model has become a common practice to improve forecasting accuracy [24], [25], and a combination of several models is employed in traffic flow forecasting [26]–[29].

The moving average (MA) and exponential smoothing (ES) models are popular in time series forecasting, and their strength lies in their good short-term accuracy combined with quick low-cost updating. However, the MA and ES models do not handle trend or seasonality well [30]. The autoregressive MA (ARIMA) model and artificial NNs are often compared, with mixed conclusions regarding their forecasting performance. The ARIMA model and the Box–Jenkins methodology are quite flexible in several typical time series such as the pure autoregressive (AR), pure MA, and combined AR and MA (ARIMA) models. The major limitation of the ARIMA model is the preassumed linear form of the model. Artificial NNs were introduced as efficient tools for modeling and forecasting approximately two decades ago. The major advantage of NNs is their flexible nonlinear modeling capability. Zhang [24] pointed out that neither the ARIMA model nor NNs are adequate to model and forecast time series because the ARIMA model cannot deal with nonlinear relationships, and the NN model alone is not able to handle linear and nonlinear patterns equally well.

It is desirable to exploit the strengths of each individual approach, which should, in turn, produce a better overall result. In this paper, a data aggregation (DA) approach for traffic flow forecasting is presented that is based on an NN. The objective of DA is to maximize useful information content by combining data and knowledge from different models.

II. Data Structure and Aggregation Strategy

In this section, we describe the DA approach for vehicle traffic flow forecasting. The traffic flow data are collected from certain data collection points and are aggregated in 1-h periods,

Manuscript received January 9, 2007; revised June 3, 2007 and February 14, 2008. Current version published February 27, 2009. This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project HKU 7176/07E, by the University of Hong Kong under Grant 10207394, by the National Natural Science Foundation of China under Grant 50578064 and Grant 70629001, by the Natural Science Foundation of Guangdong Province, China, under Grant 06025219, and by the National Basic Research Program of China under Grant 2006CB705500. The Associate Editor for this paper was H. Mahmassani.

M.-C. Tan and Z.-R. Guan are with the Department of Mathematics, College of Information Science and Technology, Jinan University, Guangzhou 510632, China (e-mail: tanmc@jnu.edu.cn; jameswingkwan@163.com).

S. C. Wong is with the Department of Civil Engineering, University of Hong Kong, Hong Kong (e-mail: hhecwsc@hkucc.hku.hk).

J.-M. Xu is with the College of Traffic and Communication, South China University of Technology, Guangzhou 510641, China (e-mail: aujmxu@scut.edu.cn).

P. Zhang is with the Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, Shanghai 200072, China (e-mail: pzhang@mail.shu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2008.2011693

24 h/day. Let $q(t)$ be the 1-h traffic flow that is collected within the time interval $(t-1, t]$ (or t for short), where t is an integer. $q(t)$ is the source time series. By analyzing the observed traffic flow data, it can be found that the traffic flow pattern is almost cyclical every week and that it is similar every weekday (Monday to Friday) and similar every weekend (Saturday and Sunday). Thus, three relevant time series are constructed for the DA approach. They are the daily similarity time series $s_1(t)$, the weekly similarity time series $s_2(t)$, and the hourly time series $s_3(t)$.

1) $s_1(t)$ is a set that includes the previous traffic flow record within the same time interval on k_1 days before $q(t)$

$$s_1(t) = \{q(t-24k_1), q(t-24(k_1-1)), \ldots, q(t-24)\}.$$

2) $s_2(t)$ is a set that includes the traffic flow record in sequential k_2 weeks before $q(t)$. The data in time series $s_2(t)$ will be on the same weekday or weekend

$$s_2(t) = \{q(t-7\times 24\times k_2), q(t-7\times 24\times (k_2-1)) \ldots, q(t-7\times 24)\}.$$

To forecast the traffic flow at time interval t on a certain Thursday, for example, we also select the data at time interval t on the previous k_2 Thursday.

3) $s_3(t)$ is a set that includes the previous k_3 traffic flow data before $q(t)$

$$s_3(t) = \{q(t-k_3), q(t-k_3+1), \ldots, q(t-1)\}.$$

Different models can be selected to forecast the three relevant time series. Let $\hat{q}_i(t)$ be the forecast value that results from model i for time series $s_i(t)$, $i = 1, 2, 3$.

In the DA stage, an NN model is used to produce the final predictions

$$\hat{q}_{\mathrm{DA}}(t) = f(\hat{q}_1(t), \hat{q}_2(t), \hat{q}_3(t)) \tag{1}$$

where $f(.)$ is the nonlinear function that is determined by the trained NN.

There are many popular models, including the naïve, MA, ES, nonseasonal ARIMA, and seasonal ARIMA (SARIMA) models, that can be applied to time series prediction [16], [24]. Choosing a proper model to forecast each of the time series $s_1(t)$, $s_2(t)$, and $s_3(t)$ is the primary task. Two important factors are considered in the model-selection stage: effectiveness and simplicity.

As the MA and ES models cannot handle trend or seasonality well [31] and the ARIMA model is often used to forecast hourly time series [6], [7], we use the ARIMA model to forecast time series $s_3(t)$. The MA and ES models are chosen to forecast $s_1(t)$ or $s_2(t)$. It is found in the case study that choosing the MA model for $s_1(t)$ and the ES model for $s_2(t)$ or exchanging them does not significantly affect the final prediction, and there is no evidence that either method is superior.

A notable characteristic of traffic flow is that it shows a very repeatable pattern in time. Time series that exhibit a repeatable pattern are modeled through the use of seasonal differencing and seasonal parameters. Such models are called a SARIMA model. We try to replace the nonseasonal ARIMA model with the SARIMA model to forecast $s_3(t)$. It is found that using the SARIMA model to forecast $s_3(t)$ produces a better prediction than using the nonseasonal ARIMA model, but the final prediction accuracy of the DA model shows no improvement in many cases with the replacement of the nonseasonal ARIMA model with the SARIMA model. This is because the daily similarity of time series $s_1(t)$ and the weekly similarity of time series $s_2(t)$ have already captured the repeatable pattern of traffic flow, and consequently, the trained NNs in the DA stage have the ability to reveal the repeatable pattern.

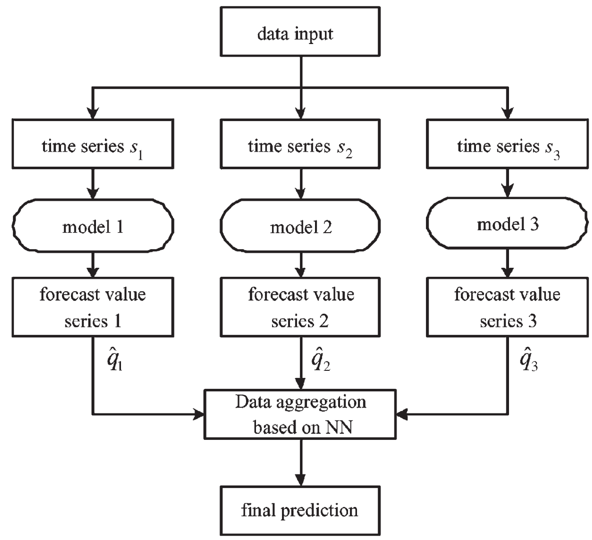


Fig. 1. Framework of the DA approach.

The naïve model has the simplest form and is used to serve as the worst case approach [16]. We use the naïve and SARIMA models for comparison purposes in the case study.

Finally, the process to produce the forecast value $\hat{q}_{\mathrm{DA}}(t)$ by the DA approach is summarized as follows (see Fig. 1).

Step 1) Select time series $s_1(t)$, and use the MA model to produce the forecast value $\hat{q}_1(t)$ (see Section III-A1).

Step 2) Select time series $s_2(t)$, and use the ES model to produce the forecast value $\hat{q}_2(t)$ (see Section III-A2).

Step 3) Select time series $s_3(t)$, and use the nonseasonal ARIMA model to produce the forecast value $\hat{q}_3(t)$ (see Section III-A3).

Step 4) In the DA stage, an NN model is used to produce the final forecast value by (1) (see Section III-B).

III. Formulation of DA for Time Series Forecasting

In this section, y_t denotes the actual value at period t, $\hat{y}_{t+1}$ is the forecast value for the next period, and $\hat{y}_{t+p}$ is the forecast for p periods into the future.

A. Individual Submodels

1) MA Model: An MA of order k is computed by

$$\hat{y}_{t+1} = \frac{y_t + y_{t-1} + y_{t-2} + \cdots + y_{t-k+1}}{k} \tag{2}$$

where k is the number of terms in the MA [31].

The MA technique deals only with the latest k periods of known data; the number of data points in each average does not change as time continues. The MA model does not handle trend or seasonality well.

2) ES Model: ES is a forecasting method that seeks to isolate trends or seasonality from irregular variation. It has been found to be most effective when the components that describe the time series change slowly over time [32].

Holt developed an ES method, Holt's two-parameter method [31], which allows for evolving local linear trends in a time series

$$H_t = \alpha y_t + (1-\alpha)(H_{t-1} + b_{t-1}) \tag{3}$$

$$b_t = \gamma(H_t - H_{t-1}) + (1-\gamma)b_{t-1} \tag{4}$$

$$\hat{y}_{t+p} = H_t + b_t p \tag{5}$$

where H_t is the new smoothed value, α is the smoothing constant for the data $(0 \leq \alpha \leq 1)$, γ is the smoothing constant for the trend estimate $0 \leq \gamma \leq 1$, b_t is the trend estimate, and p is the period to be forecast into the future. The weights can be selected by minimizing the measure of the mean square error (MSE).

3) ARIMA Model: A general ARIMA model of order (r, d, s) representing the time series can be written as

$$\phi(B)\nabla^d y_t = \theta(B)e_t \tag{6}$$

where e_t represents the random error term at time t, B is a backward-shift operator defined by $By_t = y_{t-1}$ and related to ∇ by $\nabla = 1 - B$, $\nabla^d = (1-B)^d$, and d is the order of differencing. $\phi(B)$ and $\theta(B)$ are the AR and MA operators of orders r and s, respectively, which are defined as

$$\phi(B) = 1 - \phi_1 B - \phi_2 B - \cdots - \phi_r B^r \tag{7}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B - \cdots - \theta_s B^s \tag{8}$$

where $\phi_i (i = 1, 2, \ldots, r)$ are the AR coefficients, and $\theta_j (j = 1, 2, \ldots, s)$ are the MA coefficients.

Box and Jenkins [33] developed a practical approach to building ARIMA models that has had a fundamental impact on time series analysis and forecasting applications. The Box–Jenkins methodology includes model identification, parameter estimation, diagnostic checking, and model forecasting [24] and consists of the following three iterative steps. First, we determine whether the time series is stationary or nonstationary. If it is nonstationary, it is transformed into a stationary time series by applying a suitable degree of differencing. This gives the value of d. Then, appropriate values of r and s are found by examining the autocorrelation function (ACF) and partial ACF (PACF) of the time series. Having determined r, d, and s, the coefficients of the AR and MA terms are estimated using a nonlinear least squares method. In this paper, the time series was analyzed using statistical software.

B. DA Based on an NN

The NN model that is employed in this paper consists of an input layer with three neurons, a hidden layer with n_1 neurons, and an output layer with one neuron (see Fig. 2). The NN model maps the input vector $(\hat{q}_1(t), \hat{q}_2(t), \hat{q}_3(t))$ to the output vector $\hat{q}_{\mathrm{DA}}(t)$, in which the hyperbolic tangent sigmoid transfer function is used in the hidden layer and the linear transfer function is used in the output layer. In the training or learning stage of the NN model, the weights or parameters of the network are iteratively modified on the basis of a set of input–output patterns known as a training set to minimize the deviance or error between the output that is obtained by the network and the observed output. The weights are initialized to small values based on the technique of Nguyen and Widrow [37]. We normalize the data to a value between 0 and 1. The number of hidden units and the learning rule are chosen through systematic experimentation. The learning rule that is commonly used in this type of network is the backpropagation (BP) algorithm or gradient descent method. Several different modifications of the BP learning algorithm are tried in the training course.

To obtain a network that is capable of generalizing and performing well with new cases, data samples are usually subdivided into three sets: 1) a training set, 2) a validation set, and 3) a test set [38]. During the learning stage of the network, an excessive number of parameters or weights in relation to the problem at hand and to the number of training data may lead to overfitting. This phenomenon occurs when the model fits the irrelevant features that are present in the training data too closely instead of fitting the underlying function that relates the inputs and outputs. This will result in the loss of the capacity to generalize learning to new cases [35].

In this paper, cross validation and an early-stopping technique are used in the optimizing training process to avoid

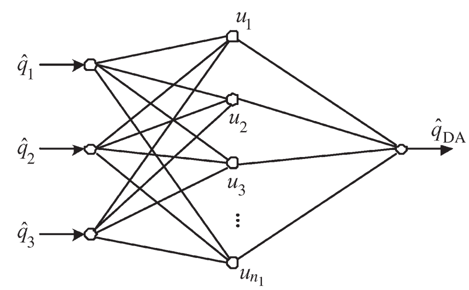


Fig. 2. Feedforward BP network.

Fig. 3. Data collection point at Xia Yuan, Guangzhou.

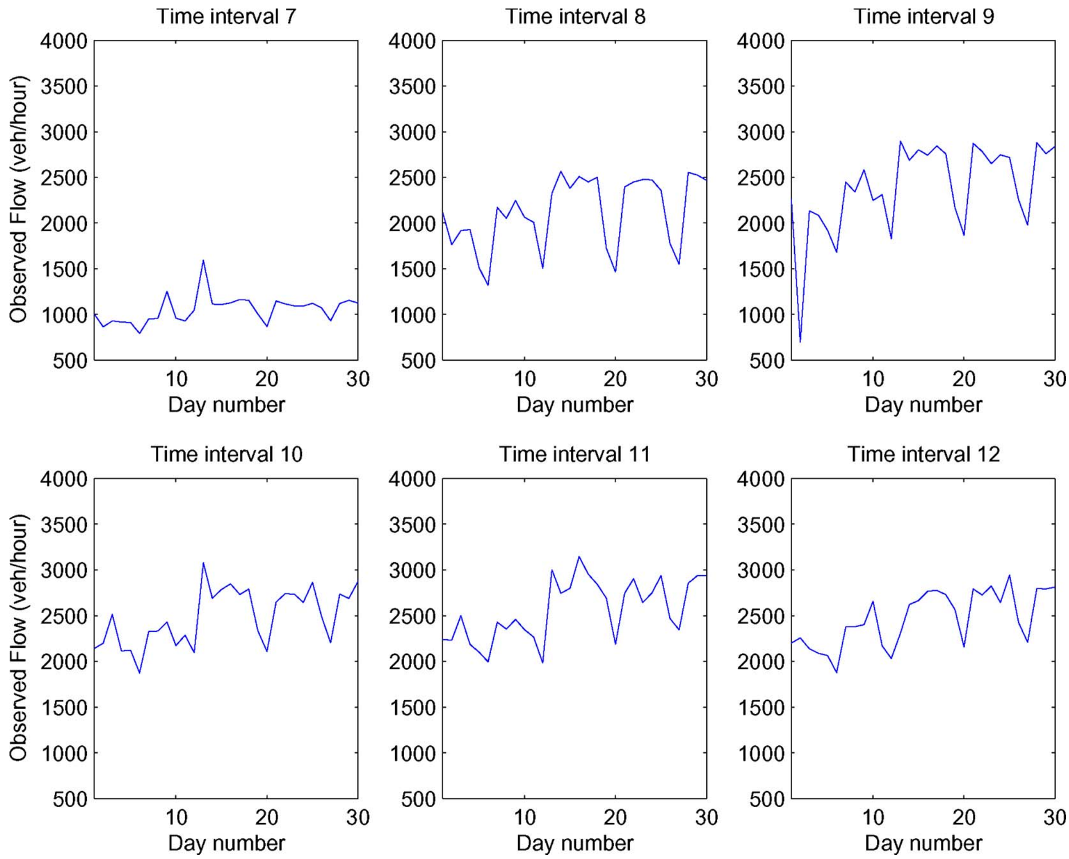


Fig. 4. Observed traffic flow at time intervals 7, 8, 9, 10, 11, and 12 in November 2005.

the overtraining (or overfitting) phenomenon of NNs. Early stopping means that the termination of training is controlled by the error of the validation sets, rather than by the error of the training set.

IV. Case Study

A. Study Area

A data set from January 1, 2005 to December 30, 2005 was collected from a detector on National Highway 107, Xia Yuan, Huangpu, Guangzhou, Guangdong, China (see Fig. 3). The traffic flow data were aggregated and averaged into 1-h periods, 24 h/day.

B. Goodness-of-Fit Statistics

Three goodness-of-fit statistics are used to assess the forecast accuracy of the results.

1) The root MSE (RMSE) is calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}. \tag{9}$$

2) The percentage absolute error (PAE) is calculated as

$$\mathrm{PAE}(n) = \frac{|y_n - \hat{y}_n|}{y_n} \times 100\%. \tag{10}$$

3) The mean absolute percentage error (MAPE) is calculated as

$$\mathrm{MAPE} = \frac{1}{N}\sum_{n=1}^{N}\frac{|y_n - \hat{y}_n|}{y_n}. \tag{11}$$

Here, y_n and $\hat{y}_n$ are the observed and the forecast values of observation n, respectively, and N is the total number of observations.

C. Selection of Models and Parameters

In the case study, we first construct the three relevant time series. Then, individual submodels are built as follows.

1) For time series $s_1(t)$, we try different orders of $1 \leq k = k_1 \leq 10$. That is, we use the traffic flow at time interval t on previous k successive days to forecast the traffic flow at time interval t on the current day. The MA model is used to forecast these time series.

For example, Fig. 4 shows the traffic flow at time intervals $7, 8, \ldots, 12$ for 30 days in November 2005. We vary the order k in (2) in the MA model. Fig. 5 shows the RMSEs of the MA models with different values of k for time interval 11, in which the RMSE is calculated based on $30 - k$ observations (days) because the traffic flow on the first k days in November cannot be predicated using the MA model of order k. Based on the results, we choose $k = 3$ when the RMSE is the smallest. Fig. 6 shows the observed and forecast traffic flows at time interval 11 for $k = 3$ in the MA model.

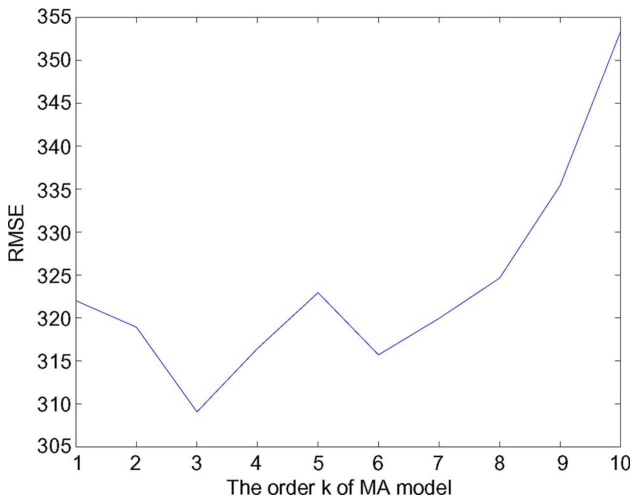


Fig. 5. RMSEs of the MA models varying with order k for $s_1(11)$.

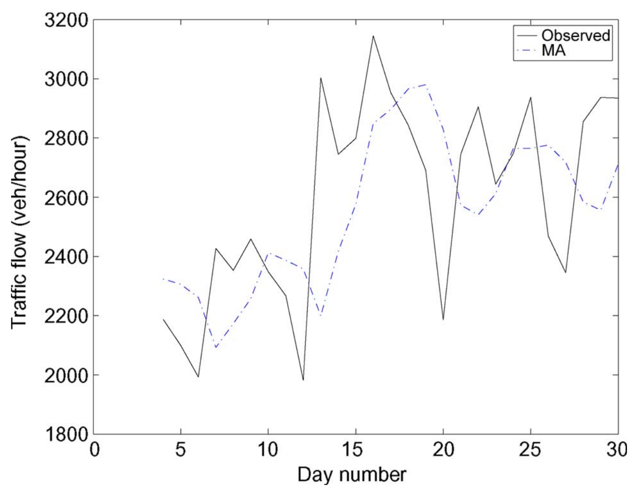


Fig. 6. Observed traffic flow and predictions that result from the MA model for $s_1(11)$.

2) For time series $s_2(t)$, we select as much data as possible from January 2005 forward to forecast the same time interval for the same weekday or weekend. Fig. 7 shows six series at time intervals $7, 8, \ldots, 12$ on all Tuesdays in 2005. Holt's ES method is employed to forecast these time series.

For example, to forecast the traffic flow at time interval 7 on Tuesday, November 29, 2005, we use the data at time interval 7 on all Tuesdays from January 5 to November 22 in 2005. Fig. 8 shows the traffic flow $s_2(7)$ at time interval 7 on all Tuesdays in 2005. To choose the proper Holt's ES model, we try all of the combinations of the smoothing constants $\alpha = 0, 0.1, \ldots, 0.9, 1$ and $\gamma = 0, 0.1, \ldots, 0.9, 1$. We select the combination ($\alpha = 0.1, \gamma = 0.1$) to fit time series $s_2(7)$ when the sum-of-square error is the smallest. The predictions that are produced by the ES method are also shown in Fig. 8.

Note that time series $s_1(t)$ and $s_2(t)$ can be generated directly from the source data. A computer program can automatically produce the predictions from the MA and ES models if the parameters are specified, such as the order $k = 3$ in the MA model and the smoothing constants $\alpha = 0.1$ and $\gamma = 0.1$ in the ES model (see Figs. 6 and 8).

3) For time series $s_3(t)$, by experimentation, we select $k_3 = 48$. That is, we use the traffic flow data in the previous 48 h (two days) to forecast the traffic flow at the current time, because Hanke *et al.* [31] have pointed out that more than 40 observations are required to develop the ARIMA model. The Box–Jenkins methodology of ARIMA modeling is employed.

We observe that time series $s_3(t)$ is nonstationary. We try to transform it into a stationary time series by applying a suitable degree of differencing d. Then, we examine the ACF and PACF of the time series to find the appropriate values of r and s. Here, we denote the ARIMA model with the parameters r, d, and s as ARIMA(r, d, s).

For example, we take 48 hourly traffic flow data on November 21 and 22, 2005. This time series is differenced once, and the differenced data vary about a fixed level, i.e., zero. After checking the ACF and PACF of the differenced data, the parameters can be chosen as (r, $s =$ 0 or 1) and $d = 1$. Table I shows the p-value and residual mean square for the three models ARIMA(1, 1, 0), ARIMA(0, 1, 1), and ARIMA(1, 1, 1).

In Table I, the $p - \text{value} = 0.809$ in model ARIMA(1, 1, 1) indicates that coefficient θ_1 is not significant from zero at the 5% level, so θ_1 can be dropped from the model. Thus, model ARIMA(1, 1, 1) is not chosen. Of the other two models, the residual mean square of the ARIMA(1, 1, 0) model is the smallest. The $p - \text{value} = 0.002$ in model ARIMA(1, 1, 0) is good enough to indicate that the coefficient ϕ_1 is significant from zero. Therefore, this model is the best fitting model and is chosen to forecast time series $s_3(t)$. The resultant model can be explicitly written as $\hat{q}_3(t) = q(t-1) + 0.4351(q(t-1) - q(t-2))$. Fig. 9 shows a comparison of the observed and predicted traffic flows from the ARIMA(1, 1, 0) model for the chosen time series.

D. Design of NNs

In designing NN models, the number of neurons in the hidden layer is an important feature that needs to be carefully chosen. To avoid the overtraining or overfitting problem of NNs, we use three data sets: 1) a training set, 2) a validation set, and 3) a testing set. Although there is no precise rule on the optimum size of these data sets, it is recommended that the training set should be the largest [38].

For example, the 2005 data $(\hat{q}_1(t), \hat{q}_2(t), \hat{q}_3(t), q(t))$ from Sunday, November 6 to Saturday, November 12 are chosen as the training set, in which $(\hat{q}_1(t), \hat{q}_2(t), \hat{q}_3(t))$ is the input to the NN and $q(t)$ is the output. The 2005 data $(\hat{q}_1(t), \hat{q}_2(t), \hat{q}_3(t), q(t))$ on November 13 and 14 are chosen as the validation set. The 2005 data on Sunday, November 20, are chosen as testing set 1 and the data on Wednesday, November 23 as testing set 2.

A series of NNs with different numbers of neurons in the hidden layer are trained. The number of neurons varies from

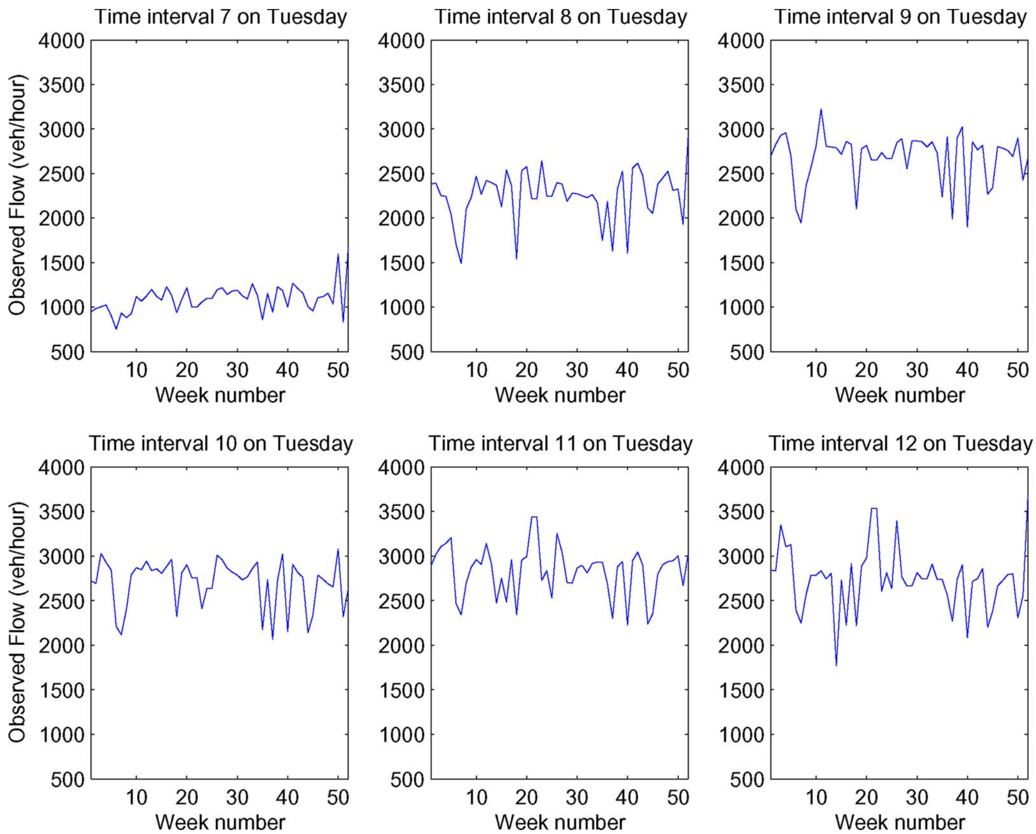


Fig. 7. Observed traffic flow on all Tuesdays in 2005.

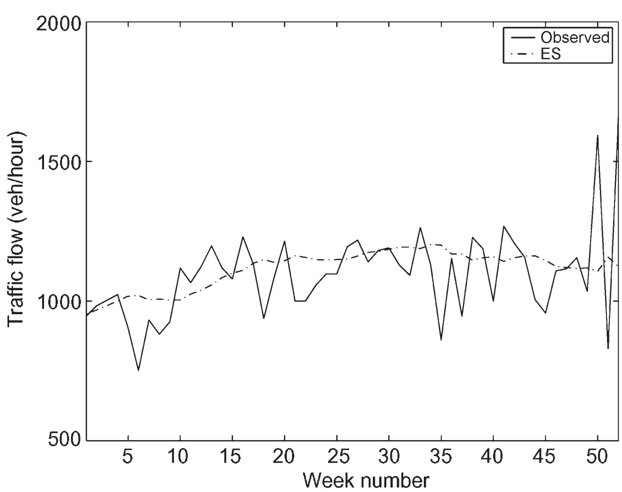


Fig. 8. Observed traffic flow and predictions that result from the ES model for $s_2(7)$ on Tuesday.

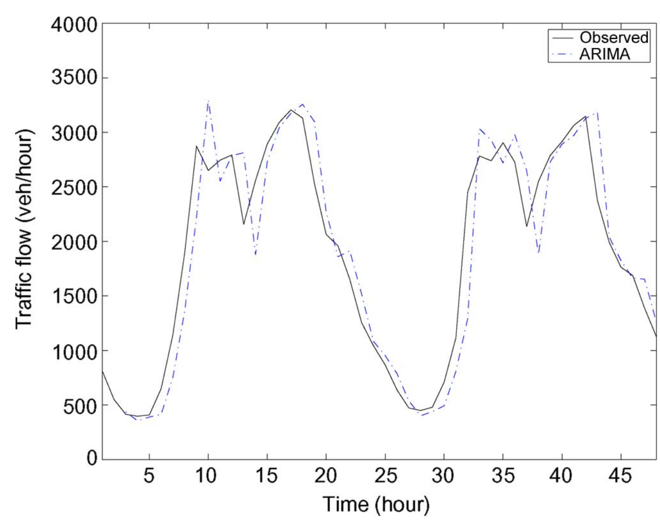


Fig. 9. Observed traffic flow and predictions that result from the ARIMA(1, 1, 0) model for $s_3(t)$.

TABLE I
RESULTS FROM DIFFERENT ARIMA MODELS

Model	Coefficients	p-value	Residual mean square
ARIMA(1,1,0)	$\phi_1 = 0.4351$	0.002	131838
ARIMA(0,1,1)	$\theta_1 = -0.4512$	0.001	132566
ARIMA(1,1,1)	$\phi_1 = 0.3664$	0.260	134708
	$\theta_1 = -0.0832$	0.809	

3 to 20, and the RMSEs are calculated for both the training and the testing set. According to its generalization ability on the testing set, the lower the value of the RMSE is, the better the network model is. Fig. 10 shows the curve of the RMSE versus the number of hidden-layer neurons when the Levenberg–Marquardt (LM) BP algorithm is used. In Fig. 10, we find that the best number of hidden-layer neurons is 12. Therefore, a 3-12-1 NN model is selected for further study.

The curves of the RMSEs for the training and validation sets versus the learning epochs that use the technique of early-stopping training are shown in Fig. 11, which shows that the RMSE swiftly decreases in both the training and validation sets when the epoch is less than eight. The RMSE achieves the lowest value when the epoch is eight and remains almost steady when the epoch increases.

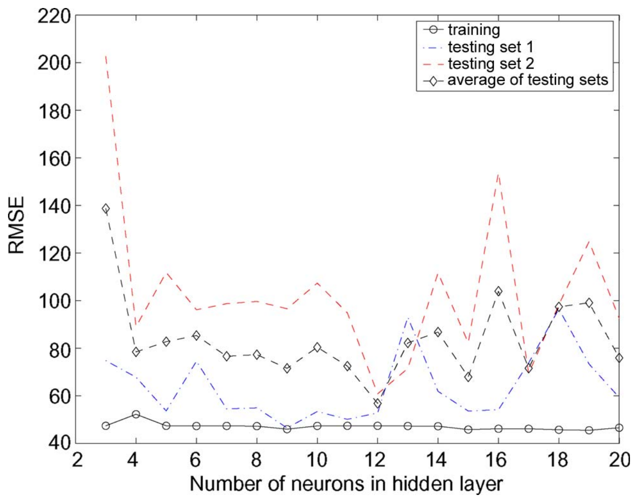


Fig. 10. Hidden-node numbers versus the MSE on the training and testing sets.

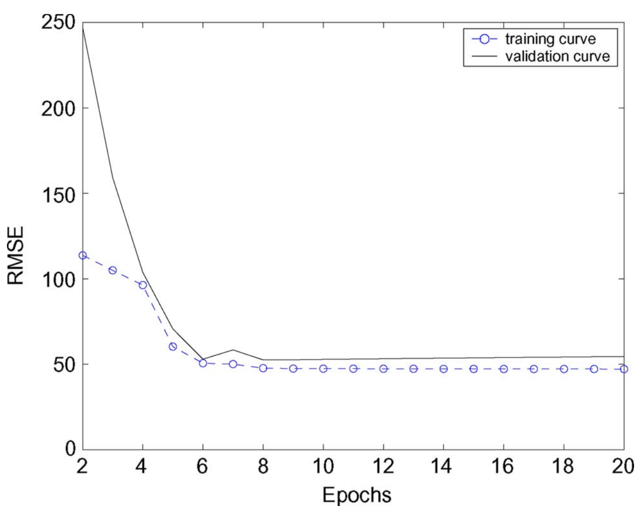


Fig. 11. RMSE of the training and testing sets versus the learning epochs.

TABLE II
LEARNING ALGORITHM COMPARISON

Learning algorithm*	Training epochs	Min MSE for training set	RMSE for test set 1	RMSE for test set 2
LM	10	0.00099	53.45	107.38
RP	4293	0.00099	54.69	104.39
GD	5000	0.00336	325.42	428.96
GDA	5000	0.00176	241.37	170.23
GDM	5000	0.00388	257.12	117.08
GDX	5000	0.00142	216.34	193.06

*LM: Levenberg-Marquardt backpropagation; RP: resilient backpropagation; GD: gradient descent backpropagation; GDA: gradient descent with adaptive learning rate back-propagation; GDM: gradient descent with momentum backpropagation; GDX: gradient descent with momentum and adaptive learning rate backpropagation.

Several BP learning rules are selected in the training course. Let the maximum training epoch be 5000 and the MSE convergence goal be 0.001. The learning epochs and the RMSEs of the outputs of the trained NN on testing sets 1 and 2 are listed in Table II, in which the early-stopping technique is not applied. Table II shows that the convergence goal is met at epoch 10 by the LM algorithm and at epoch 4293 by the resilient backpropagation (RP) algorithm, but the maximum training epoch is reached and the convergence goal is not met by the other algorithms. The RMSEs by the LM algorithm is similar to that by the RP algorithm, while the LM algorithm has the quickest convergence. The LM algorithm is the best learning rule in this case. Similar studies have been carried out in traffic flow forecasting on other data sets. It is found that the number of neurons in the hidden layers of the NN is between 4 and 24.

E. Comparison of Results

The DA approach, denoted by the DA model, integrates three submodels (MA, ES, and ARIMA) using an NN. The predictions from the MA, ES, and ARIMA models are used as input to the NN in the DA stage, and the output of the trained NN is the final prediction.

Several single-source models, including the naïve, ARIMA, NP, and NN models, are applied to time series $s_3(t)$. The ARIMA model is the same as that in the DA approach. We compare their predictions on the testing sets with those of the DA model in the case study.

1) The naïve (or no-change) model for traffic flow forecasting has the simplest form

$$\hat{q}_{\mathrm{Na}}(t) = q(t-1)$$

where $\hat{q}_{\mathrm{Na}}(t)$ is the forecast value at time interval t.

2) NP is a potential approach to traffic flow forecasting [16], [36]. Kernel estimation is a nonparametric estimation technique, and the applicability of this technique requires that both the kernel function and bandwidth be suitably chosen [15]. Smith *et al.* [16] provided an extensive comparison of the forecasting performance of six non-parametric methods, which are called "straight average," "weighted by inverse of distance," "adjusted by $V(t)$," "adjusted by both $V(t)$ and $V_{\mathrm{hist}}(t)$," "adjusted by both weight and distance," and so forth. It is demonstrated that among these models, the "adjusted by $V(t)$" model has the simplest form and relatively satisfactory performance, i.e.,

$$\hat{V}(t+1) = \frac{1}{K}\sum_{i=1}^{K}\frac{V_i(t+1)V_c(t)}{V_i(t)}$$

where $V(t)$ is the traffic flow at time interval t, $V_c(t)$ is an element in the current state, and $V_{\mathrm{hist}}(t)$ is the historical average volume at the time of day and day of the week that are associated with time interval t (see [16] for a detailed description of the model). In this paper, forecasts are generated using K-nearest neighbor forecasts for values of K between 5 and 40, and it is found that when $K = 25$, the model produces the best prediction on the basis of MAPEs.

3) An artificial NN as a single-source model is used for comparison with the DA approach. To show the difference, the NN in the DA approach is denoted by NN1 and the NN for the purpose of comparison is denoted by NN2.

As described above, the NN1 model is trained to fit the nonlinear function in (1). The NN2 model is used to fit the nonlinear relationship

$$\hat{q}_{\mathrm{NN}}(t) = f_1\left(q(t-1), q(t-2), \ldots, q(t-l)\right).$$

The NN1 and NN2 models have different inputs. The inputs of the NN1 model are the predictions from the MA, ES, and ARIMA models. The inputs of the NN2 model are the traffic flow records at previous l successive time intervals, and its output is the prediction of traffic flow at time interval t.

The NN2 model is constructed with one input layer, one hidden layer, and one output. The number of inputs l and the number of hidden neurons in the NN2 model are also optimized by experimentation.

Note that all of the models (naïve, ARIMA, NP, and NN2) are employed to produce forecasts for time series $s_3(t)$. To produce the prediction of $s_3(t)$, each model may need a different length of historical data. Hence, the sample size should be properly chosen for each model. For example, the length of time of the training data for the naïve model is no more than one day before the forecasting time, while that for the training data for the NP model covers several weeks before. For each model, we choose the parameters by observing the best fitting or forecasting. We compare their forecasts on the same test sets.

Testing set 1 is the traffic record on Sunday, November 20, 2005. Testing set 2 is the traffic record on Wednesday, November 23, 2005. The training sample for the above models covers the data from January 1, 2005 to November 12, 2005. The validation set for the NN2 model is the observed traffic flow record on November 13 and 14, 2005.

The models above and the DA model are evaluated for their out-of-sample forecasting performance using a recursive training sample (see [39] for a detailed discussion on this method). Let T be the forecasting origin, and we generate forecasts for time periods $T+1, T+2, \ldots, T+N$. The procedure of the N-step-ahead forecast is given as follows.

1) Select a training set. Let T_0^* $(< T)$ be the index of the final traffic flow record in the training set.
2) Identify and estimate each model (naïve, ARIMA, NP, and NN2) using the training set.
3) Identify and estimate the models (MS, ES, ARIMA, and NN1) in the DA approach, where the ARIMA model is the same as that in step 2), using the methods that are described in Sections IV-B, C, and D. In this step, we use the historical data before T that cover the training set in step 1), as we may need much more data to construct time series $s_1(t)$ and $s_2(t)$.
4) Compute the N-step-ahead forecasts for the models (naïve, ARIMA, NP, NN2, and DA).
5) Advance the time index by one and increase the training sample by one, i.e., set $T_1^* = T+1$, and go to step 4). Estimate the same models and iterate over all of the values of the testing data set.
6) Compute the MAPEs between the N-step-ahead predictors and the observed values.

As we advance the time index, we do not identify the parameters for the models (naïve, ARIMA, NP, NN2, and DA). We use the same model parameters that are identified using the training sample. The advantages and disadvantages of this evaluation procedure have been discussed in [40]. When evaluating the DA model, the N-step-ahead forecast may need the one-step-ahead forecasts from time series $s_1(t)$ and $s_2(t)$. For example, to produce the two-step-ahead forecast $\hat{q}_{\mathrm{DA}}(t+2)$ by the DA model, we need $(\hat{q}_1(t+2), \hat{q}_2(t+2), \hat{q}_3(t+2))$, in which $\hat{q}_1(t+2)$ is the one-step-ahead forecast from time series $s_1(t+2)$, $\hat{q}_2(t+2)$ is the one-step-ahead forecast from time series $s_2(t+2)$, and $\hat{q}_3(t+2)$ is the two-step-ahead forecast from time series $s_3(t)$.

TABLE III
MAPEs (IN PERCENTAGE) (IN VEHICLES PER HOUR) OF MODELS WITH DIFFERENT TIME HORIZONS (ON TESTING SET 1)

Forecast model	One-hour ahead	Two-hour ahead	Three-hour ahead
Naïve	16.39	30.78	44.29
ARIMA	12.56	22.14	33.27
NP	9.78	14.56	**11.85**
NN2	10.63	19.21	23.15
DA	**5.92**	**8.44**	12.21

TABLE IV
MAPEs (IN PERCENTAGE) (IN VEHICLES PER HOUR) OF MODELS WITH DIFFERENT TIME HORIZONS (ON TESTING SET 2)

Forecast model	One-hour ahead	Two-hour ahead	Three-hour ahead
Naïve	18.98	33.93	48.94
ARIMA	12.63	23.42	33.45
NP	9.07	**10.06**	12.25
NN2	9.21	20.09	22.46
DA	**6.84**	10.22	**11.30**

Tables III and IV show the MAPEs of the two testing sets with the different forecast horizons $N = 1, 2,$ or 3. In this paper, the NN2 model has three inputs, the number of neurons in the hidden layer is 16, and the LM algorithm is used for its training. The bold numbers in the tables indicate the best performance in each column.

The MAPE statistics in Tables III and IV show that the predictions that result from the NP and DA models are better than the predictions that result from the naïve, ARIMA, and NN2 models. The predictions that result from the NN2 model are better than the predictions that result from the ARIMA model. The DA approach provides the best one-step-ahead forecasting results compared with the other models. The NP and DA models have almost the same performance in two-step-ahead and three-step-ahead forecasting. Although ARIMA models are quite flexible in many time series, the results are not ideal when the time series is highly nonlinear. As the NN1 model in the DA approach has one input from the ARIMA model, the poor performance of the ARIMA model affects the prediction accuracy of the DA model. Fig. 12 shows the PAEs that result from the NN1 model in the DA approach on the training, validation, and testing sets. The PAE is not always small in the application of the DA approach because of the highly nonlinear characteristic of traffic flow.

V. CONCLUSION

Combining different modeling schemes for the improvement of prediction capability has become a common practice in

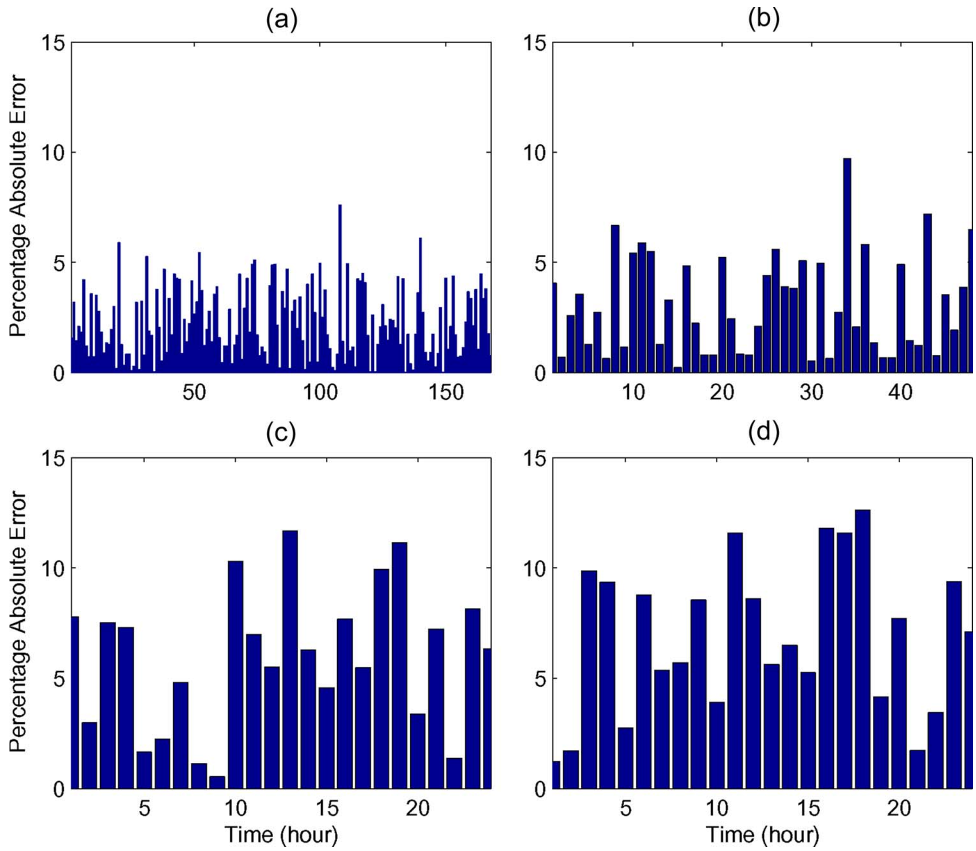


Fig. 12. PAEs of the DA model for the training, validation, and testing sets.

different application fields, but the approaches vary, and thus, worthy of investigation are specific applications. In this paper, aiming at the repeatable pattern of traffic flow, three relevant time series have been constructed, and three models have been used to forecast these series. The DA strategy that has been proposed uses NN technology and aggregates the forecasting values that result from the multiple models. The weekly similarity time series and daily similarity time series can be easily constructed from the source time series, and the forecasting value of the MA and ES models can be automatically obtained by a computer program once the parameters of the models are specified. The predictions of the hourly time series by the ARIMA model can easily be obtained with statistical software. Therefore, the proposed DA approach is not a time-consuming but, rather, a feasible job.

The relevant time series make full use of the information in the source time series that is collected on a single detector. By analyzing the forecasting performance of the naïve, ARIMA, NP, NN, and DA models, we have shown that the DA model can provide results that are more accurate than those of the other models.

If an accident or nonrecurring congestion happens, the repeatable pattern of traffic flow is lost, and the prediction accuracy will be affected. In such a case, one solution is to install sufficient detectors on the upstream to track the traffic flow. How to use the DA approach in a multiple-detector environment deserves further study. The DA method that is based on NNs may have considerable potential in forecasting technology.

ACKNOWLEDGMENT

The authors would like to thank the four anonymous referees for their helpful suggestions and critical and constructive comments on an earlier version of the paper.

REFERENCES

[1] B. Abdulhai, H. Porwal, and W. Recker, "Short-term traffic flow prediction using neuro-genetic algorithms," *J. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 3–41, Jan. 2002.
[2] M. C. Tan, J. M. Xu, and Z. Y. Mao, "Traffic flow modeling and on-ramp optimal control in freeway," *Chin. J. Highw. Transp.*, vol. 13, no. 4, pp. 83–85, 2000.
[3] M. C. Tan, C. O. Tong, and J. M. Xu, "Study and implementation of a decision support system for urban mass transit service planning," *J. Inf. Technol. Manage.*, vol. XV, no. 1/2, pp. 14–32, 2004.
[4] M. C. Tan, C. O. Tong, S. C. Wong, and J. M. Xu, "An algorithm for finding reasonable paths in transit networks," *J. Adv. Transp.*, vol. 41, no. 3, pp. 285–305, 2007.
[5] M. C. Tan, L. B. Feng, and J. M. Xu, "Traffic flow prediction based on hybrid ARIMA and ANN model," *Chin. J. Highw. Transp.*, vol. 20, no. 4, pp. 118–121, 2007.
[6] B. M. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling," *Transp. Res. Rec.*, vol. 1776, pp. 194–200, 2001.
[7] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.
[8] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–131, Mar. 2006.
[9] M. S. Dougherty and M. R. Cobbett, "Short-term inter-urban traffic forecasts using neural networks," *Int. J. Forecast.*, vol. 13, no. 1, pp. 21–31, Mar. 1997.

[10] C. Ledoux, "An urban traffic flow model integrating neural networks," *Transp. Res., Part C–Emerg. Technol.*, vol. 5, no. 5, pp. 287–300, Oct. 1997.
[11] D. Shmueli, "Applications of neural networks in transportation planning," *Prog. Plann.*, vol. 50, no. 3, pp. 141–204, Oct. 1998.
[12] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *Eur. J. Oper. Res.*, vol. 131, no. 2, pp. 253–261, Jun. 2001.
[13] H. B. Yin, S. C. Wong, J. M. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res., Part C–Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, Apr. 2002.
[14] L. W. Lan and Y. C. Huang, "A rolling-trained fuzzy neural network approach for freeway incident detection," *Transportmetrica*, vol. 2, no. 1, pp. 11–29, 2006.
[15] N.-E. El Faouzi, "Nonparametric traffic flow prediction using kernel estimator," in *Proc. 13th ISTTT*, J.-B. Lesort, Ed., Lyon, France, 1996, pp. 41–54.
[16] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res., Part C–Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.
[17] H. Chen and S. Grant-Muller, "Use of sequential learning for short-term traffic flow forecasting," *Transp. Res., Part C–Emerg. Technol.*, vol. 9, no. 5, pp. 319–336, Oct. 2001.
[18] G. Huisken, "Soft-computing techniques applied to short-term traffic flow forecasting," *Syst. Anal. Model. Simul.*, vol. 43, no. 2, pp. 165–173, Feb. 2003.
[19] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res., Part C–Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, Jun. 2005.
[20] H. A. Edgerton and L. E. Kolbe, "The method of minimum variation for the combination of criteria," *Psychometrika*, vol. 1, no. 3, pp. 183–187, Sep. 1936.
[21] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Oper. Res. Q.*, vol. 20, pp. 451–468, 1969.
[22] M. J. Lawerence, R. H. Edmundson, and M. J. O'Connor, "The accuracy of combining judgmental and statistical forecasts," *Manag. Sci.*, vol. 32, no. 12, pp. 1521–1532, 1986.
[23] S. Makridakis,"Why combining works?" *Int. J. Forecast.*, vol. 5, no. 4, pp. 601–603, 1989.
[24] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomput.*, vol. 50, pp. 159–175, Jan. 2003.
[25] Y. Gao and M. J. Er, "NARMAX time series model prediction: Feedforward and recurrent fuzzy neural network approaches," *Fuzzy Sets Syst.*, vol. 150, no. 2, pp. 331–350, Mar. 2005.
[26] Y. Jiang, "Prediction of freeway traffic flows using Kalman predictor in combination with time series," *Transp. Q.*, vol. 57, no. 2, pp. 99–118, 2003.
[27] N. E. El Faouzi, "Combining predictive schemes in short-term traffic forecasting," in *Proc. 14th ISTTT*, A. Ceder, Ed., Jerusalem, Israel, 1999, pp. 471–487.
[28] W. Zheng, D. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *J. Transp. Eng.*, vol. 132, no. 2, pp. 114–121, Feb. 2006.
[29] M. V. D. Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res., Part C–Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, Oct. 1996.
[30] B. L. Bowerman and R. T. O'Connell, *Forecasting and Time Series: An Applied Approach*, 3rd ed. Florence, Italy: Brooks/Cole, 1993.
[31] J. E. Hanke, A. G. Reitsch, and D. W. Wichern, *Business Forecasting*, 7th ed. Boston, MA: Prentice–Hall, 2001.
[32] R. A. Yaffee and M. McGee, *Introduction to Time Series Analysis and Forecasting*. San Diego, CA: Academic, 2000.
[33] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, 2nd ed. San Francisco, CA: Holden-Day, 1970.
[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, pp. 318–362.
[35] E. B. Baum and D. Haussler, "What size net gives valid generalization?" *Neural Comput.*, vol. 1, no. 1, pp. 151–160, 1989.
[36] E. Matzner-Løbber, A. Gannoun, and J. G. De Gooijer, "Nonparametric forecasting: A comparison of three kernel-based methods," *Commun. Stat., Theory Methods*, vol. 27, no. 7, pp. 1593–1617, 1998.
[37] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proc. Int. Joint Conf. Neural Netw.*, 1990, vol. 3, pp. 21–26.
[38] C. M. Bishop, *Neural Network for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
[39] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: An analysis and review," *Int. J. Forecast.*, vol. 16, no. 4, pp. 437–450, Oct. 2000.
[40] D. D. Thomakos and J. B. Guerard, Jr., "Naïve, ARIMA, nonparametric, transfer function and VAR models: A comparison of forecasting performance," *Int. J. Forecast.*, vol. 20, no. 1, pp. 53–67, Jan.–Mar. 2004.

Man-Chun Tan received the M.S. degree in mathematics from Jinan University, Guangzhou, China, in 1995 and the Ph.D. degree in control theory and control engineering from the South China University of Technology, Guangzhou, in 2000.

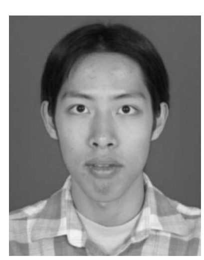
From July 2000 to September 2000, he was a Research Assistant with the Department of Civil Engineering, University of Hong Kong, Hong Kong. From June 2001 to May 2002, he was a Research Associate with the University of Hong Kong. He is currently a Professor with the Department of Mathematics, College of Information Science and Technology, Jinan University. His research interests include traffic flow prediction and control, transportation planning, stability analysis of dynamic systems, and artificial neural networks.

S. C. Wong received the B.Sc.(Eng) and M.Phil. degrees from the University of Hong Kong, Hong Kong, and the Ph.D. degree from University College London, London, U.K.

He is currently a Professor with the Department of Civil Engineering, University of Hong Kong. His research interests include optimization of traffic signal settings, continuum modeling for traffic equilibrium problems, land use and transportation problems, dynamic highway and transit assignment problems, urban taxi services, and road safety.

Jian-Min Xu received the Ph.D. degree from the South China University of Technology, Guangzhou, China.

He is currently a Professor and a Doctoral Advisor with the College of Traffic and Communication, South China University of Technology. As a Project Principal, he has accomplished many national or ministerial/provincial projects, some of which are National 863 High-Tech Programs or National Fund Projects. He is the author or a coauthor of more than 100 journal papers. He is the holder of three patents. His research interests include intelligent transportation systems and intelligent control theory and its application.



Zhan-Rong Guan received the B.S. degree in management information system from the College of Information Science and Technology, Jinan University, Guangzhou, China, in July 2005. From September 2005 to June 2007, he worked towards the M.S. degree in applied mathematics with the Department of Mathematics, Jinan University.

His research interests include traffic flow forecasting and design of information systems.

Peng Zhang received the B.Sc. degree from Sichuan University, Sichuan, China, and the M.Sc. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China.

He is currently a Professor with the Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, Shanghai, China. His research interests include traffic flow theory and computational methods.